教育研究與發展期刊　第二十一卷第三期　2025 年 9 月　頁 1-36

DOI　10.6925/SCJ.202509_21(3).0001

「循證決策」與「數據驅動教學」概念在教育領域之實踐：以日本學力調查為例

郭晏輔　國立臺灣師範大學教育學系博士後研究員
陳佩英　國立臺灣師範大學教育學系優聘教授
李哲迪　國立臺灣師範大學科學教育中心副研究員

摘　要

在資訊科技發展與績效責任制等多重因素影響下，教育領域中「循證決策」（EBPM）與「數據驅動教學」（DDI）等概念日漸盛行，然而目前較缺少關於實務案例的討論。本文以日本「全國學力・學習狀況調查」與各自治體層級獨立實施的學力調查為例，說明這些學力評估如何成為資料、數據、證據的生產來源，供政策制定者、學者、學校領導者、教師等行動者進行參考，透過意義與資訊的生產，形塑政策制定與教學行動上的參照。本文透過日本官方文件、書籍、媒體等資料之整理，彙整出日本各地應用學力調查的實際案例，並將其連結學術概念。在數據資料和人工智慧蓬勃發展的趨勢下，本研究可供臺灣教育工作者思考如何更有效善用數據，以改善教學行動。

關鍵詞：循證決策、數據驅動教學、全國學力・學習狀況調查、國家大型調查、日本教育

【通訊作者】郭晏輔　電子郵件：latwony5227@hotmail.com

The Implementation of Evidence-Based Policy Making and Data-Driven Instruction in Education: A Case Study of Academic Assessments in Japan

Yen-Fu Kuo
Postdoctoral Researcher, Department of Education, National Taiwan Normal University
Pei-Ying Chen
Distinguished Professor, Department of Education, National Taiwan Normal University
Che-Di Lee
Associate Researcher, Science Education Center, National Taiwan Normal University

Abstract

Driven by multiple factors including advancements in information technology and performance accountability systems, concepts such as Evidence-Based Policy Making (EBPM) and Data-Driven Instruction (DDI) have gained increasing prominence in the field of education. Nevertheless, there remains a relative lack of discussion on practical case studies. This study examines Japan's National Assessment of Academic Ability along with various locally administered academic assessments to illustrate how these evaluations serve as sources for generating data, information, and evidence. These outputs, in turn, inform decision-making processes among policymakers, researchers, school leaders, and teachers, shaping both policy formulation and instructional practices through the production of meaning and information.

Drawing on official Japanese government documents, academic literature, and media reports, this study compiles practical examples of how different regions in Japan utilize academic assessments and contextualizes them within relevant academic frameworks. Against the backdrop of rapid developments in data analytics and artificial intelligence, this research provides insights for Taiwanese educators on how to more effectively leverage data to enhance instructional practices.

Keywords: Evidence-Based Policy Making (EBPM), Data-Driven Instruction (DDI), National Assessment of Academic Ability, National large-scale student assessment, Japanese Education

Yen-Fu Kuo's E-mail: latwony5227@hotmail.com (Corresponding Author)

壹、研究背景

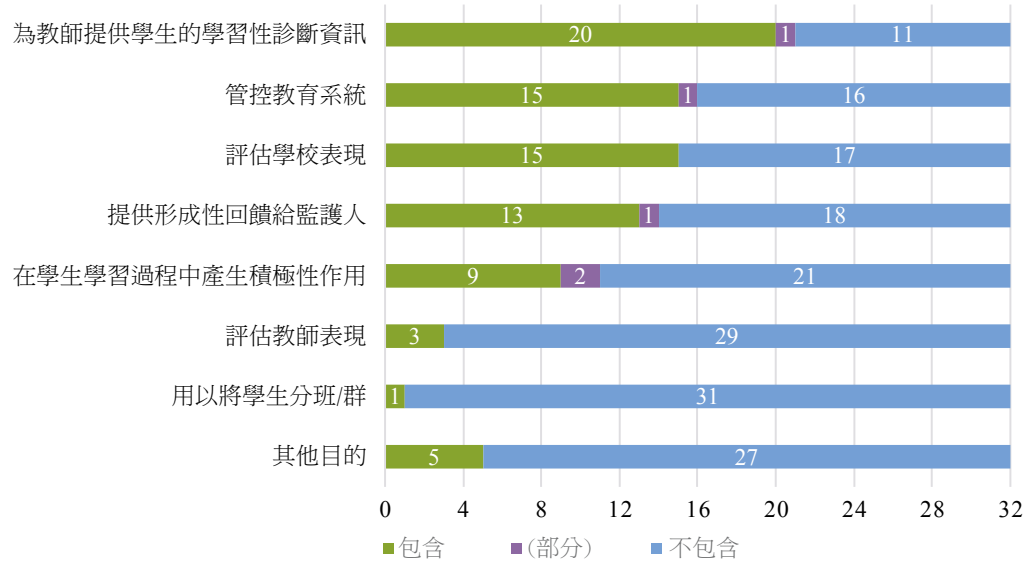
隨著教育市場化與績效責任制的推波助瀾，各國政府愈加關注教育投入與產出之間的關係。為了讓教育預算最大效益化，政府部門日益重視「循證決策」（evidence based policy making, EBPM）原則，精準評估教育投入與產出之間的關係（Lingard et al., 2015），如何發展具備科學性的教育指標並且應用之，便成為教育部門積極發展的方向。特別在 PISA 等國際大型調查（International large scale assessments, ILSAs）實施下，此種趨勢愈加明顯，各國希冀將學力評估作為一種政策槓桿（policy lever），以利數據治理與教育改革之實踐（Teodoro, 2022; Volante & Mattei, 2024）。

除了運用 ILSAs，諸多國家也研製「國家大型調查」（National large-scale student assessment, NALSA），藉此種標準化評估來檢視教育概況（Lee et al., 2023），如由美國國家教育統計中心主導之「全國教育進展評估」（National Assessment of Educational Progress, NAEP），用以協助各州推進教育改革，此後在「不讓任何孩子落後」（No Child Left Behind, NCLB）的法案實施中，NAEP 更凸顯其教育指標實施的重要性（Lee & Reeves, 2012）。而在大型調查的推波助瀾下，澳洲於本世紀初期開展之「全國評估計畫 ── 識字與算數」（National Assessment Program - Literacy and Numeracy, NAPLAN），被期望透過問責制與教育表現的透明性來促進學校品質提升（Thompson, 2013）。根據經濟合作暨發展組織（Organisation for Economic Co-operation and Development, OECD）（2023）報告指出，目前有 32 個國家實行 NALSA，包含加拿大的「泛加拿大評估計畫」（Pan-Canadian Assessment Program）、德國的 VERA（Vergleichsarbeiten in der Schule）等。有別於「測驗、考試」（tests/examinations）作為一種高風險（high-stake）檢定，將會影響學生生涯路徑，具備「回應性」的「評量」（assessments）則關注教學指導（Morozumi & Tanaka, 2023）。在 OECD（2023）報告中列舉了這些全國性評量的功能，如圖 1 所示。

根據圖 1 所示，NALSA 的功能可分別針對學生、教師與學校、教育決策單位三個方面進行回饋，其功能對應了 EBPM 以及「數據驅動」（data-driven）概念在教育場域的應用，而本文的核心即是在探究具體的實踐方法。本文以日本於 2007 年開始實施的「全國學力・學力狀況調查」（日文：全国学力・学習状況調査，官方英譯：National Assessment of Academic Ability，以下簡稱「全國學調」），以及

圖 1

全國性評量主要用途分類



註：修改自 OECD（2023, p. 407）。

各自治體獨立實施之學力調查為例（若不特別區分，本文則以「日本學調」代稱兩種類型調查），說明日本各層級教育部門如何善用評量工具，作為教育政策與教學方案制定過程中的參照物。

值得注意的是，日本學調能作為證據、數據的代表工具，鑲嵌於日本的時代脈絡。2019 年文部科學省提出之「第三期教育振興基本計畫」即言明希冀依循 EBPM 原則並結合 PDCA（Plan-Do-Check-Act cycle）的管理，藉由跨領域教育專才協作與各式指標之訂立來促成教育政策之推進與學校管理，其中特別提及利用社會科學方法於規劃全國學調，並偕同大學與研究單位的合作，研擬可提供全國、各自治體層級整體教育改善發展之方法（文部科学省，2019a）。因應此趨勢，國立教育政策研究所（National Institute for Educational Policy Research, NIER）於 2021 年成立「教育資料科學中心」（教育データサイエンスセンター），該組織主要目的即在協助中央、地方部門教育性質資料之處理，範疇包含全國學調與各式 ILSAs，希冀發揮資料驅動與證據為本的功效（国立教育政策研究所，2025）。此外，近年文部科學省推動之「教育 DX」（Digital Transformation）與「GIGA 學校」

（Global and Innovation Gateway for ALL）重大教育政策變革，均再再顯示日本教育界對於 EBPM 與數據驅動之重視。

若依照 OECD 報告的標準，我國目前尚缺少一個嚴格定義上的 NALSA，或是目前正處於發展階段。雖然如會考、大學入學考試等高風險測驗存在已久，但此類測驗無法提供實質上的教學回饋，或是結合學生背景擴大政策規劃。而雖各地方政府有參與與規劃局部地域性的評量，如科技化學習扶助等，雖然一定程度上能夠給予學校或教學者回饋，但無法進行全國規模性的比較。另外，我國也積極參與各式國際評比（如 PISA、TIMSS），雖數據上可以提供政策回饋，但是施測周期長，無法發揮即時之效。種種因素限制了 EBPM 在臺教育推動上的順暢性。因而，本文希冀運用日本學調作為借鏡，透過官方資料、日文書籍與媒體檔案作為素材的蒐集整理，希冀提供臺灣未來在在評量或是相關資料數據庫上開發與應用之建議。

貳、文獻探討

一、教育中的 EBPM 與數據驅動

資訊科技的發展促使了數據累積的速度，此脈絡下，各國政府更為重視證據在公共政策上的應用，臺灣也不例外，近期如國家發展委員會委託政治大學辦理之《循證決策集刊》中即可發現，數據與證據的應用在當前產、官、學方面日漸盛行。

循證決策概念之萌生與奉理性為圭臬之現代主義社會的發展息息相關，特別在日益複雜的社會中，如何有效地解決治理上的問題，將成為證據運用的核心價值（Howlett, 2009）。此處的證據一詞，可類比為經由科學方法所產出之研究，因此，許多探討 EBPM 的論述將重心專注在「研究」與「政策」之間的連結（Nutley et al., 2007）。然而，對於否認存在純粹客觀科學事實的後現代主義論者，大多質疑 EBPM 的主張過於理想（Howlett, 2009）；反之，EBPM 的擁護者則認為，若能參酌更多的資訊量，將有助於有效落實政策（Nilsson et al., 2008）。

將 EBPM 概念應用在教育的風氣明顯較晚於公共衛生、醫療和公行領域，部分原因在於教育政策與實踐往往仰賴於主觀經驗或是意識形態。

> 在大多數社會中，教育被要求做愈來愈多的事情，標準愈來愈高，也被賦予更大的責任，但卻只分配到有限的資源。教育的議程（agenda）往往由政治意

> 識形態、傳統智慧、民間傳說和一廂情願的想法所驅動，以滿足經濟、商業、雇主、法律與秩序、公民社會、家長選擇，並至少在口頭上滿足學習群體中兒童、年輕人和成人的需求和利益。（Davies, 1999）

Davies（1999）認為這將導致「希望戰勝理性，情感與直覺戰勝證據」的狀況，他也指出，伴隨教育改革而生的新方案時常缺乏有效性驗證，部分原因在於教育實踐鮮少使用（準）實驗設計，長期追縱等具有高品質的研究來進行佐證。同一時期 Hargreaves（2007）透過醫學與教育兩個領域之比較，認為後者呈現「非累積性」（non-cumulative）的弱勢，不僅令教育研究零散化，且缺乏對於教學現場的支持。因此，EBPM 被期待能夠導入教育場域中，透過科學化的方式精進教育實踐。

不過在探討 EBPM 的實踐前，應對「證據」一詞進行琢磨，根據日本內閣官房（2022）所釋出之「EBPM 指導手冊」（EBPM ガイドブック）中，詳細說明了如何參酌證據以利政策落實之法，也包含了對於證據一詞的界定。

表 1
證據的類型與實例

種類			內容	針對弱勢階層子女的支援方案之具體例子
證據（廣義）	資料、事實	資料、事實	統計資料、公聽會等能夠協助掌握現狀的資訊	1. 弱勢群體子女的比例變化狀況 2. 其他經濟狀況下的子女之升學率變化狀況 3. 對貧困家庭的生活狀況進行意見聽取
		未來預測	現況持續下對於未來的預測	1. 子女數量的未來預測 2. 處於弱勢家庭的子女數之未來預測數量
	證據（狹義）	量化因果效果	透過統計方式所揭示的量化政策因果效果	1. 通過隨機對照實驗揭示的教育計畫效果 2. 生活習慣良好的兒童和生活習慣不良的兒童的學力比較
		質性因果效果	透過質性方式所揭示的政策因果效果	基於焦點小組訪談揭示的生活習慣改善計畫的效果

註：修改自內閣官房（2022，頁 23）。

根據表 1，證據可分為廣義與狹義兩類，廣義範疇包含了一般性資料與現況評估，狹義的證據則指涉透過科學方法產出的研究或調查。換言之，「證據」是一種集合概念。又如「證據金字塔」所界定（如圖 2、表 2），證據的價值會因生產方式來區分，儘管愈高層級證據具備更高的信效度，然也因生產成本較高而呈稀少性。

圖 2
證據金字塔示意與說明

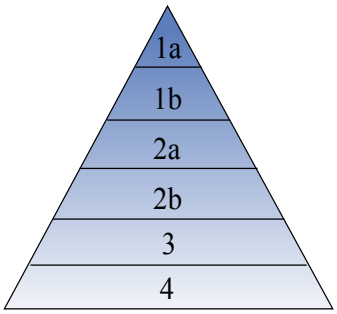


表 2
證據金字塔示意與說明

等級	內容
1a	系統性回顧、後設分析
1b	隨機分派實驗（RCT）
2a	差異中差異（DID）、斷點迴歸（RDD）
2b	迴歸分析、世代研究（Cohort study）
3	平均數檢定、相關分析、質性報告
4	專家與實務者的意見

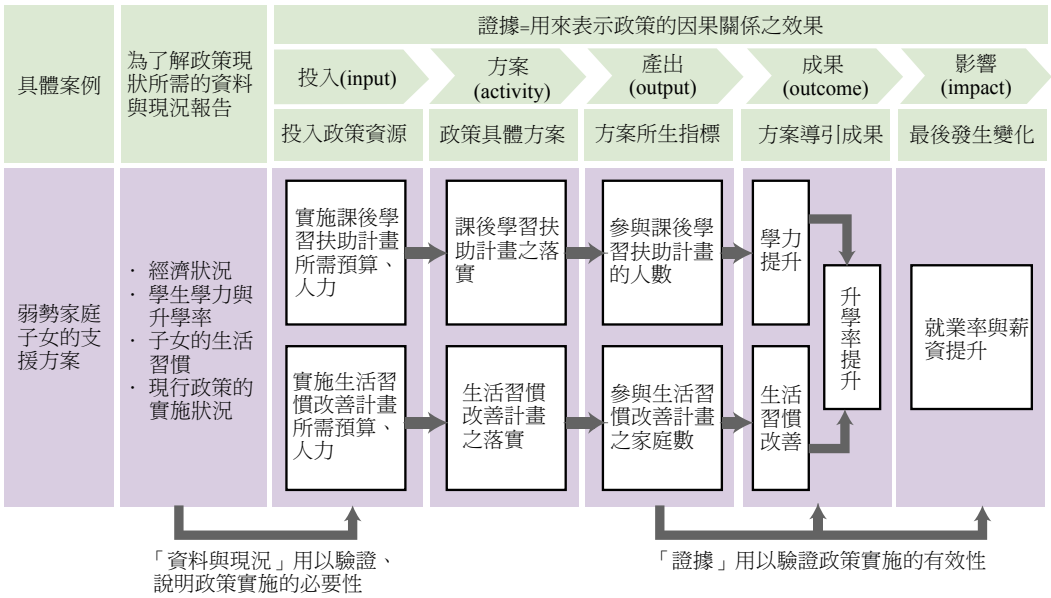
註：修改自內閣府（2020，頁 10）。

完成對於證據的基本瞭解，下一步需探討證據如何進入決策。就宏觀而言，證據應用型態多樣，Weiss（1979）界定學術研究證據應用於決策過程的七種形式，包含：（一）知識導向模型（Knowledge Driven Model）；（二）問題解決模型（Problem-Solving Model）；（三）交互作用模型（Interactive Model）；（四）政治性模型（Political Model）；（五）策略性模型（Tactical Model）；（六）啟蒙性模型（Enlightenment Model）；（七）研究作為社會智識事業的一環（Research as Part of the Intellectual Enterprise of Society）。這七種模型的前兩者屬於理想情況，實際上證據很少被直接應用於決策，多數情況下證據是採「間接」的方式漸進改變決策者的思維。另一方面，基於政治性模型的基礎，證據時常被用以驗證當前政策實施的有效性（Sanderson, 2002）。具體案例上，如指導手冊所示（圖 3），證據能被應用於製作「邏輯模型」（logic model），當中，廣義證據能夠鋪陳政策實施的必要性，包含現況評估與議題發現，而狹義的高階證據則體現在圖右半部，希望藉因果關係研究，用以檢證政策的有效性。

由此見得，EBPM 取向的政策規劃重視具體指標的建構與可計算性的成果，並藉資源投入與實際成效的連結來呈現因果關係。然而，大部分對於 EBPM 的誤解在於期待單一或是複數的證據、研究能規劃一套完善的政策，事實上政策實施的良窳涉及多種因素，無法完全依賴科學研究產出的科學證據（Howlett, 2009），此外，學界與政策制定者間缺乏完善溝通機制也是導致 EBPM 無法順利推行的原因（Sonobe & Kurosaki, 2024）。

圖 3

循證決策之邏輯模型實踐模組



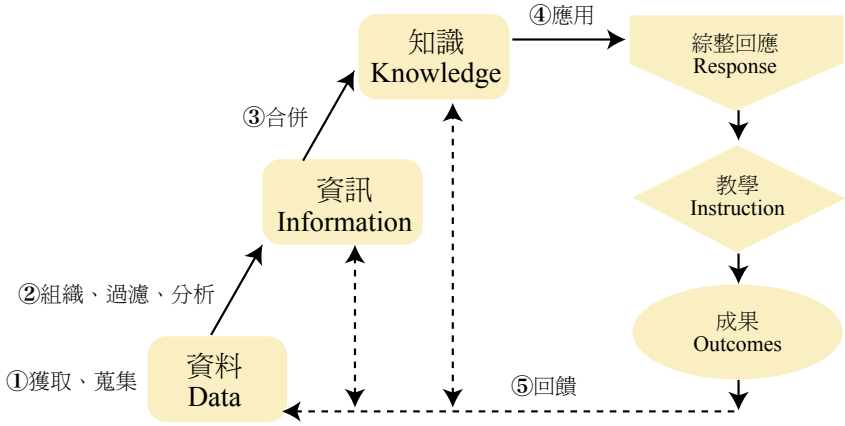
註：修改自內閣官房（2022，頁 24）。

除了 EBPM，近期也有相關學術概念指涉將數據結合進入教育場域中的行動計畫，如同「數據驅動決策」（Data Driven Decision Making, DDDM）（Kaspi & Venkatraman, 2023; Lai & Schildkamp, 2013）、「數據驅動教學」（Data Driven Instruction, DDI）（Bambrick-Santoyo, 2010）、「數據素養」（Data Literacy）在教育上的應用（Filderman et al., 2022）。這些概念相互之間存在異同，但基本的原理與知識管理中的 DIKW（data-information-knowledge-wisdom）金字塔相似（Rowley, 2007），如圖 4 呈現，數據驅動在教育上的應用主要探討如何有效蒐集「數據」（Data），並進行分析形成「資訊」（Information），教育工作者再從這些訊息中萃取，規劃，形成「知識」（Knowledge），這些知識得以用以回應教學需求，之後，透過教學成果來回饋到各種資料應用環節（Mandinach et al., 2006; Marsh & Farrell, 2015）。

就具體行動來說。Lai 與 Schildkamp（2013）將教學現場的資料定義為「一種透過收集、組織而來的資訊，反映了學校的各個層面」，其包含質性資料，如訪談、參訪記錄，也包含量化資料，如測驗結果和相關圖表，然也有學者認為只有可計算

圖 4

數據驅動概念應用於教育之模組



註：修改自 Marsh 與 Farrell（2015, p. 272）。

的數據才屬於資料。而關於資料應用，學者見解分歧，如 Schildkamp（2019）指出，資料能夠以一種迭代（iterative）的方式應用於教育組織，學校在進行校務改革時都需設定出清晰的目標，為達到目標，需要蒐集資料，當中包含（非）正式調查（隨機的課堂觀察或是討論）以及教育研究證據、文獻。資料蒐集完的下一步為意義建構環節，此一過程重視數據的解讀，透過專家的詮釋，數據資料可轉為知識，進而形成實踐方案，此為行動與評估的環節。Poortman 與 Schildkamp（2016）則認為資料應用於校務改善的三個主要部分是課程（如改進課程的一致性）、評量（如應用形成性評量來取代總結性評量）、與教學（為風險學生提供額外的輔導）。除了這些較為直接的用途外，資料也能以「概念」的方式呈現，藉此影響教師或是學校領導者的思維（Farley-Ripple et al., 2018; Schildkamp & Visscher, 2010）。

綜整而言，EBPM 與 DDDM、DDI 等學術概念互有重疊之處，皆講求善用客觀、科學的研究或是數據資料來規劃行動方針，然而 EBPM 更加聚焦在政策規劃與檢核面向，屬於宏觀層級，而資料驅動的意涵則較偏重於教學現場的執行，屬於微觀層級。而在數位科技發展的時代下，教育的數據、資料的種類和數量增長迅速，為了聚焦，本研究以日本的學力調查（含全國與地方層級）為例，探討 EBPM 與 DDI 的具體方案。之所以以日本做為借鏡，理由有二，其一為教育文化因素，基於同為亞洲，重視考試制度的國家，日本比起歐美諸國可能對於臺灣更有參考價值，且臺灣曾接受過日本殖民，教育型態上有所相似之處；其次為資料豐富度，日本學

調實行至今已近 20 年，過程中已累積豐富的經驗、文件、研究，深具學習價值。

二、日本學力調查的發展脈絡與實施簡介

文部科學省透過每年頒布的全國學調報告書，公告評比的實施要項，以令和 6 年（2024）的成果報告為例，全國學調的施測目的如下：

> 為了確保義務教育的機會均等，並維持和提高其水平，瞭解和分析全國範圍內的學生學習情況和學力，檢驗教育政策的成果和問題，並加以改進。同時，這些措施也有助於落實學校對學生的學習指導並改善其學習狀況。此外，通過這些方式，來建立起教育方面的持續性檢驗和改進循環。（文部科学省，2024a，頁 2）

除了學科調查，全國學調為了分析影響學習之因素，同時編製背景問卷，當中包含學習動機、方法等調查，此外也向校方調查學校資源等訊息。實施方式主要採普查方式執行，以 2024 年為例，學科調查於 4 月 18 日執行，參與的學校含公、國、私立，當中包含國小 18,703 和國中 9,665 校，學生數分別為 978,409 與 910,938 人。以下針對學調之發展背景、施測細項進行更詳細說明。

（一）實施背景

日本全國學調之實施可追溯於 1950 年代，戰後復甦的日本蔓延著對於學力平庸化的擔憂，當時在美國援助下，日本的「新教育運動」深受杜威經驗主義與進步主義之影響，提倡「問題導向學習」（高橋徹，2022）。教育界雖然贊同個體為本的教育模式，但也存在學科能力下滑之疑慮，包含無法識別基本漢字或算術，此擔憂催生學力調查之實施。1956 年，文部省以針對全國 6、9、12 年級學生抽樣，執行首次全國學力調查，科目包含國語、算數、理科、社會等，甚至包含音樂、家政、體育等領域。然在實施過程中遭遇不少反對聲浪與問題，例如「日本教職員組合」批評，學力調查違反「教育基本法」的規範，亦即國家機關不應介入教師教學內容；此外也連帶質疑政府會強化對於教師評鑑的控制力（川口俊明，2020）。其二則與學力調查掀起自治體間惡性成績競爭有關，如香川、愛媛兩縣為了提高學習調查表現，迫使成績較為落後的學生請假，或是教師公然在考試期間洩漏答案（浦岸英雄，2010）。此類問題層出不窮，導致學調於 1966 年廢止。

在前學調廢除後約 40 年全國學調再度開展，或有人稱之「全國學力調查的復

活」（久永隆一，2023），起因主要受到兩個事件之影響，分別為「學力低下爭論」（学力低下論争）與國際評比之結果。受美蘇冷戰影響，日本教育部於上世紀 60 年代逐步開展培養學生科學、數理方面的知識，「教育內容現代化」的論述逐漸興起。1971 年，以「現代化課程」為名的學習指導要領開始實施（今井康晴，2008），大幅擴充課程內容與提高難度，使得學生難以適應，故有人稱之「填鴨式教育」（詰め込み教育）。由於學習品質下降，導致出現大量校園暴力、輟學等社會問題，為因應之，日本中央教育審議會於 1975 年後開始提倡「寬鬆教育」（ゆとり教育），要旨為減少學習內容，降低授課時數，並且提高選修課程數量等面向（間渕泰尚，2019）。就如同進步主義觀所帶來的負面影響，寬鬆教育再度引發學力下滑的爭議，然而當時缺乏實證來驗證此種擔憂是否為真（松下佳代，2017）。

21 世紀伊始，由 OECD 主導的 PISA 於 2000 年首次實施，時隔三年之後，日本發生「PISA Shock」事件（Ninomiya, 2019; Takayama, 2008）。意旨為，相比第一屆的結果，日本在 2003 年的表現中，大多素養指標在世界排名出現下滑現象，該事件不僅被用以主張寬鬆教育的失敗，更引發一系列教育變革，學者松下佳代（2014）將其區分為「政策轉換的直接影響」與「構造改變的間接影響」，前者指涉學習指導要領的方針由「寬鬆教育」轉向「學力向上」，後者則關注由中央機關主導教育績效之評估。因應教育改革需求，文部科學省於 2007 年正式實施全國學調。綜整而言，川口俊明（2020）分析全國學調的實施有以下原因與目的：

1. 競爭導向的氛圍

全國學調透過公布各自治體的平均答對率，促進地區間的學習與競爭。2004 年文部科學大臣中山成彬認為，教育應強調競爭與切磋精神（望田研吾，2005）。這反映了政策從寬鬆教育轉向「學力重視」。

2. 公共教育品質的檢核

在財政赤字下，政府更加關注教育投資的效率。全國學調成為衡量教育成效的工具，並基於 EBPM 邏輯評估投入與產出，以確保公共資源運用透明。

3. 應用於教學指導

除了評估效益，全國學調結果也用於學生輔導。數據交付各層級教育機構，使學校能依學生學習狀況提供適切支援。

4. 評估學習指導要領的實施

全國學調關注教育政策與學習目標的落實，評估學生是否掌握學習指導要領，並檢視教師與教育委員會的執行情況。試題設計對應指導要領的核心概念，稍後將

詳細說明。

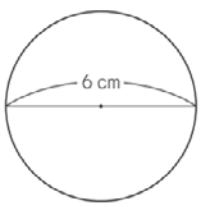
（二）實施要項

2003 年受到 PISA 衝擊後，文部科學省旋即設立 PISA・TIMSS 對應工作小組（PISA・TIMSS 対応ワーキング・グループ），小組由大學教授與官員組成，針對日本學生如何提升國際評比之表現進行研究，會中研擬出需要提高「PISA 型閱讀能力」的方針（遠藤貴広，2010），「素養」（literacy）的概念正式引入學習指導要領。爾後全國學調的題型設計上，在國語、數學科皆分為 A、B 兩種試題，當中 A 試題偏重於「知識」，B 試題則關注「活用」（小杉聡，2023；松下佳代，2014），此類似臺灣基本學科能力和素養導向學習的兩種知識分類。

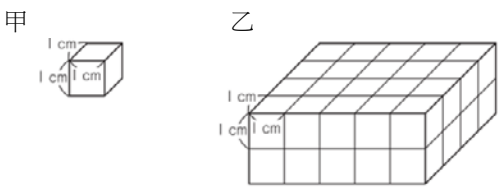
圖 5

全國學調算數 A、B 試題範例

(1) 請計算下方圓的圓周長，圓周率為 3.14。

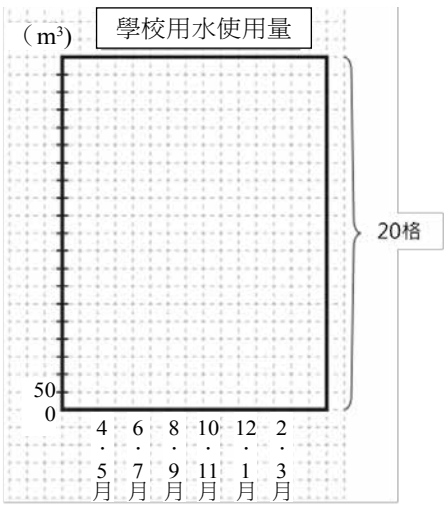


(2) 如下圖所示，使用甲圖的立方體，製作成乙圖的長方體，請問乙圖的長方體體積為何？



（A）

小明想要根據表單中每月學校用水量的資訊畫成柱狀圖，於是他在筆記本上畫了一個縱向 20 格橫向 15 格的框架，並且將縱向的每一個刻度設定為 50 立方公尺。



這樣一來，4 月和 5 月的柱狀圖可以放入框內，但是發現有幾個月的柱狀圖無法放入框內。請從下面的 1 到 5 中選擇一個月並寫下其編號。此外，請用文字和數字說明為什麼該月的柱狀圖無法放入框內。

（B）

註：修改自国立教育政策研究所（2014）。

圖 5 左為 A 試題，內容偏重基本算術能力，而圖右為 B 試題，題幹不僅較長，也重視數學知識於生活情境中的應用。由此可見，全國學調能回應日本當時課程綱要之變革，發揮政策擴散之效。自 2007 年首次實施以來，調查科目與對象選擇方式並非固定，如表 3 所示，2012 年開始每三年實施一次理科，2019 年開始每三年實施一次英文。而須留意的是在 2019 年開始試題整合化，原因與學習指導要領再度修訂下，更為強調基礎知識與活用技能的整合性有關。

表 3
全國學調歷屆施測科目

施測年分	施測方式	施測對象與施測科目	
		六年級	九年級
2007	普查	國語、算數	國語、數學
2008	普查	國語、算數	國語、數學
2009	普查	國語、算數	國語、數學
2010	抽樣	國語、算數	國語、數學
2011	抽樣	國語、算數	國語、數學
2012	抽樣	國語、算數、理科	國語、數學、理科
2013	普查	國語、算數	國語、數學
2014	普查	國語、算數	國語、數學
2015	普查	國語、算數、理科	國語、數學、理科
2016	普查	國語、算數	國語、數學
2017	普查	國語、算數	國語、數學
2018	普查	國語、算數、理科	國語、數學、理科
（以上年份之國語、算數、數學都會區分 A、B 兩類型）			
2019	普查	國語、算數	國語、數學、英語
2020	普查	國語、算數	國語、數學
2021	普查	國語、算數	國語、數學
2022	普查	國語、算數、理科	國語、數學、理科
2023	普查	國語、算數	國語、數學、英語
2024	普查	國語、數學	國語、數學

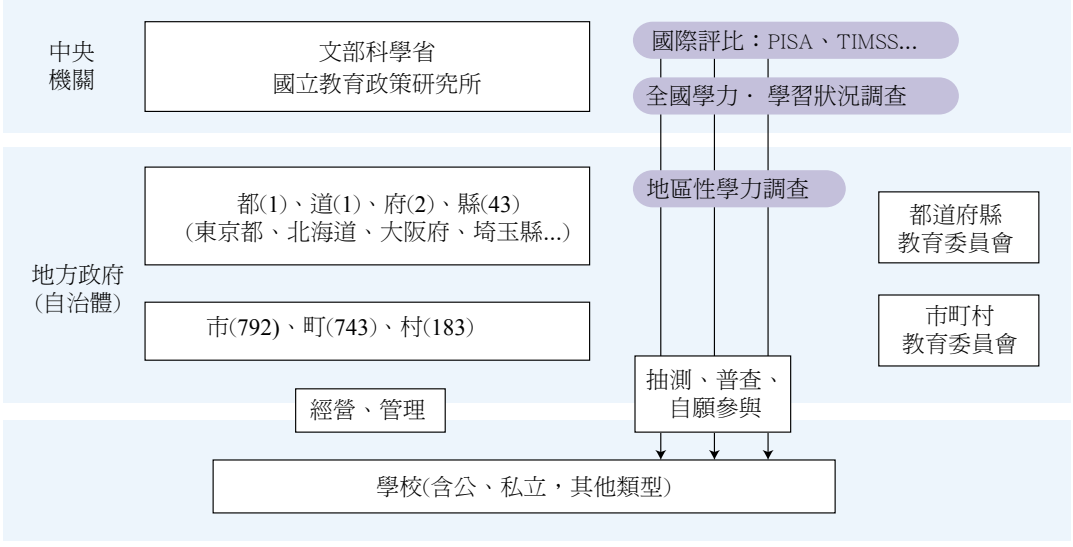
而為了搭配日本自 2019 年開始推行之 GIGA 學校數位科技導入教育政策，全國學調的評量方式將傳統紙本評量轉化為電腦化測驗（computer based testing, CBT），並結合試題反應理論（item response theory, IRT），以順應數位資訊教育趨勢，此種新型態的調查方式將於 2025 年開始局部試行（文部科学省，2024b）。

（三）地方性學力調查

日本實施的學力評量類型，不僅用於全國層級，亦包含地方自治體獨立實施的學力調查。藉圖 6 可方便理解在中央、地方政府層級架構下不同學力調查的主職機構。

圖 6

日本各行政層級施行學力評比之階層架構



在最上層中央機關部分，文部科學省下轄的國立教育政策研究所負責主持國際評比與全國學調，而在自治體層級又可大致區分為都道府縣及其下轄之市町村層級，他們大多會獨立實行地方層級學力評量，且實施方式因地而異，茲將文部科学省（2018）之調查進行翻譯並整理如表 4。

表 4
各自治體實施學力評量之概要

自治體	施測年級	施測學科	時程（月）	施測方式	是否公布	公布層級
北海道	無進行施測					
青森縣	5	國、算、理、社	8	普查	是	市、郡
岩手縣	5	國、算、理、社	10	普查	是	都、道、府、縣
宮城縣	僅對 5 年級學生進行背景問卷調查					
秋田縣	4	國、算、理	12	普查	是	都、道、府、縣
	5、6	國、算、理、社			是	
山形縣	5	特殊類型題目	4	普查	是	都、道、府、縣
福島縣	4、5、6	國、算	5	普查	是	都、道、府、縣
茨城縣	3、4、5、6	國、算、理、社	1	自願	是	都、道、府、縣
栃木縣	4、5	國、算、理	4	普查	是	都、道、府、縣
群馬縣	無進行施測					
埼玉縣	4、5、6	國、算	4	普查	是	市、町、村
千葉縣	無進行施測					
東京都	5	國、算、理、社 特殊類型題目	7	普查	是	都、道、府、縣
神奈川縣	無進行施測					

註：修改自文部科学省（2018）

囿於篇幅限制，只呈現小學階段的資料，並且將範疇限縮在日本北海道、東北與關東三地區，但也可發現各地實施方式各具特色。各地學調起始執行之時間不一，其原因也和寬鬆教育和國際評比有關，寬鬆教育調降學習時間的結果本身令諸多教育學者對於學生學習成就感到擔憂，另外也因應 2003 年 PISA Shock 的結果，對於學力觀察之事日漸積極，在此之後，大部分的自治體開始實施地方學調（北野秋男，2015）。

參、研究發現

研究結果分成三個專題，首先以工具的角度出發，說明「S-P（student-problem）表」如何成為教育議題生產的工具；其二，藉埼玉縣的案例探討地方學調落實

EBPM 的可能性；最後，透過沖繩縣案例，結合參照社會的觀點說明如何利用政策學習的方法改善學力表現。

一、S-P 表之使用：工具行動者觀點

EBPM 與 DDI 的本質強調讓數據得以轉化、生產為實務工作者得以運用的資訊，由日本學者佐藤隆博所提出之 S-P 表即為具體的方法，不少自治區教育部門利用該工具進行學調分析，促進學調的可用性。

表 5
SP 表範例

座號	題號										答對題數
	8	10	5	6	1	9	4	7	3	2	
12	1	1	1	1	1	1	1	1	1	1	10
8	1	1	1	1	1	1	1	1	1	2	9
3	1	1	1	1	1	2	1	1	3	1	8
11	1	1	1	1	1	1	1	2	3	3	7
15	1	1	1	2	1	2	1	2	1	3	6
6	1	1	3	1	1	3	1	1	2	2	6
13	1	1	1	1	2	1	2	3	3	3	5
1	1	1	1	2	1	2	2	2	1	2	5
14	1	1	3	1	2	3	1	2	2	1	5
2	1	3	2	1	2	1	2	1	1	2	5
10	2	1	1	1	2	1	3	3	3	3	4
4	1	2	2	2	2	1	2	1	3	1	4
9	1	2	1	2	1	3	3	3	3	3	3
5	2	2	1	2	2	1	2	2	2	2	2
7	2	1	2	2	3	3	3	3	2	2	1
答對人數	12	11	10	9	8	8	7	6	5	4	

註：表內數字 1 代表答對，2 代表答錯，3 為未填答。修改自富山県中学校教育研究会（2024）。

S-P 表可幫助教師根據某一份試卷中學生的答題反應，歸納分析學生的學習狀況。在表 5 當中，首先依學生的答對題數由上而下排序，比方說 12 號同學全對，排在最上列。之後根據答對人數由左至右排序試題難度，如第 8 題共有 12 人答對，

排在最左行，難度較低。排序完畢之後可繪製 S 曲線（粗線）與 P 曲線（虛線）。S 曲線按學生得分的變化繪製，P 曲線按試題難度變化繪製。S 曲線越偏右，代表高成就學生更多；曲線越陡，表示學生間得分差異越大。P 曲線位置越上，代表高難度的試題越多；曲線越陡，表示試題間難度差異越大。換言之，SP 表能夠同時分析學生能力與試題性質，並提供更多學習上的資訊，比方說 13 號與 2 號同學，儘管他們的答對題數相同，但是 2 號同學的作答結果相對特殊，他無法答對簡單的試題，反而能夠完成難度較高的題目，對此，可以深究該生的作答狀況，另一方面，針對第 1 題與第 9 題，兩題答對人數相同，然而程度中間以上與以下的學生答對該題的比率相當，這意味著有重新檢視該題品質的必要性。由於 SP 表有上述功能，因此能把「證據」、「數據」轉譯為「教學問題」。

搭配學調結果，利用 S-P 表所進一步獲得的資訊，可規劃和進行不同的教學方案。例如東京足立區獨自實行的學調中，即應用 S-P 表進行延伸設計。他們將學生能力分為不同等級，並以此提供相應教學方案，譬如針對學力需要補強的學生規劃課後班，或是單獨輔導機制，以精進基本能力；而對於較高年級，或是學力中上的學子，提供更為進階的集中式課程。另一方面，S-P 表的學習診斷功能使教師得以進一步探勘學生較不擅長的領域，並據以進行補救教學（足立区教育委員会，2024）。

整體而言，S-P 表在學調資料運用的推廣過程中，彷彿一個行動者，推動著教師將「數值」轉化為「議題」，催化教學方案的產生。S-P 表也能夠成為協助教師專業成長的機制。例如，長野縣教育委員會為提高落後於全國的學調表現，提供訪校措施來改善教師教學，措施之一便是教導教師使用 S-P 表分析學生學習狀況，這和過往只提供單一課程模式，無法進行多層面檢視以回應各校差異性問題的做法不同，S-P 表的應用確實獲得成效，參與輔導的教師或有表示能夠利用 S-P 表看到具體問題，用以精準改善教學（文部科学省，2019b）。另外在新瀉縣燕市，則結合運用 S-P 表、學校資料與全國平均表現，找出學力待加強之學生，施以扶助方案（文部科学省，2019c）。整體而言，S-P 表具有將「數值」轉化為「議題」之功能，即便當今 AI 結合數位系統能夠提供更多元的認知診斷分析，S-P 表仍能憑藉視覺化資料和相對容易製作的優勢，提供教育工作者進行 EBPM 或是 DDDM 之實踐。

二、埼玉縣案例：多層次教育數據應用之典範

埼玉縣位於東京都西北部，為首都圈範疇，又被稱作為東京的衛星都市。根

據 2025 年發布之統計（埼玉県庁，2025），埼玉縣人口約為 730 多萬人，縣總生產毛額約為 23.7 兆日圓，家庭月收入約為 56 萬日幣，這些指標都在 47 個自治體中排名第五，可凸顯出該地經濟優勢性。在自治體獨自實施學調的案例中，埼玉縣引人關注，Takayama 與 Lingard（2021）即以埼玉為對象，探討日本評量科學化的歷程，並利用行動者網絡理論說明地域性學力評比如何結合學界、政治人物與教學工作者，讓證據為本與數據驅動的概念得以實踐。該地成功的經驗甚至讓 OECD 教育與技能執行長 Andreas Schleicher 前往該地造訪，此外由於實施成效矚目，埼玉的模板也被其他自治區（如：福島縣）所學習。

（一）實施概要與技術介紹

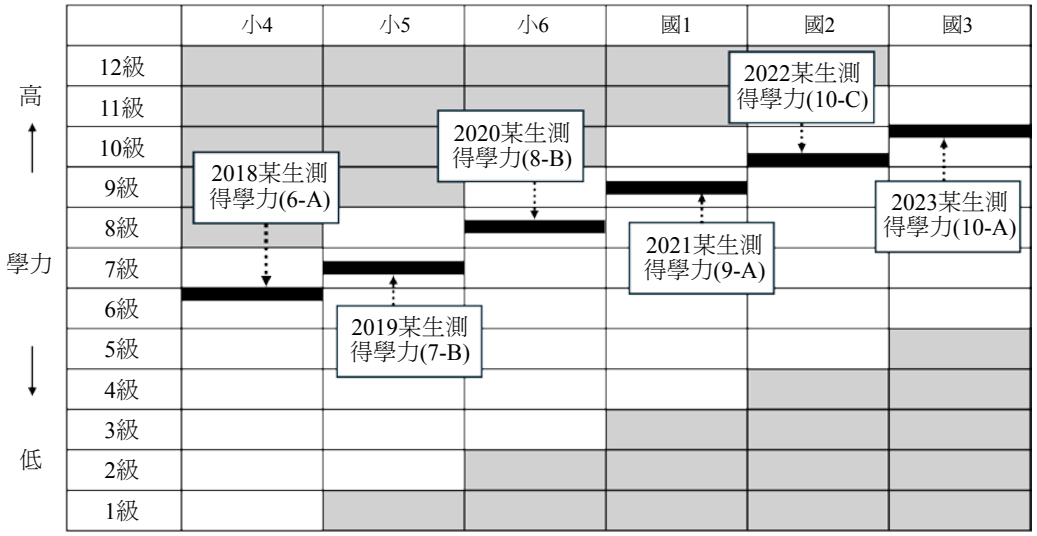
埼玉縣於 2015 年開始實施學習狀況調查（以下稱埼玉學調），調查範圍包含縣內 62 個市町村中超過 1000 所學校，由小四到國三，每年約有 30 萬學生參與此項調查，科目包含國語、數學與英文（國二開始實施）。有別於其他地區學調，埼玉學調具有三特點，分別為（1）重視縱貫性追蹤、（2）試題根據 IRT 設計、（3）進行非認知能力之調查。執行縱貫追蹤的原因與該縣設定之「第 2 期埼玉教育振興基本計畫」當中所提到「推動確保每個人都能成長的教育」（一人一人を確実に伸ばす教育の推進）有關（大根田頼尚，2019）。基於此項原則，學生參與每屆埼玉學調的資料將會被記錄保存，並且在試題編制技術上，埼玉學調運用 IRT 的「等化」（equating）技術，使得不同年級、不同屆別的試題得以在共同的量尺上進行比較。

為了能透過數值化方式清楚呈現學生學力變化樣態，埼玉學調將小四到國三，總共六個年級的學習程度區分為 12 個等級，個別年級內的試題難度橫跨七個等級（即圖 7 中白色區塊，小四的試題難度涵蓋 1 到 7 級），各個等級內又區分為 A、B、C 三個階段，某生的測驗結果會被歸類在某一個學習程度（如圖中黑色區塊）。由於學生在不同年級的測驗分數都是對應到相同的一個量尺，因此可進行比較。比方說圖 7 中黑色區塊代表某生每一年級的學力檢測結果所換算的學力等級，上升趨勢顯示該生在六年當中的學力有所成長。

透過長期追蹤調查與 IRT 所建立的量尺，便能有效評估每位學生的學業成長趨勢。由於以個體內在比較為測驗目的的核心，因此，即使某生在兩屆學業調查結果在班級排名都居於末位，只要發現其能力級數有所上升，該生學業仍會被判定為有所進步（埼玉県教育委員会，2024）。

圖 7

埼玉學調學力縱貫追縱示意圖



註：修改自埼玉県教育委員会（2024）。

（二）因果關係研究與推廣

能夠廣納不同領域學術社群的參與，並針對數據進行客觀分析，是埼玉學調能作為 EBPM 代表案例的重要原因之一。參與埼玉學調分析的學者們肯認教師對於學習成就存有關鍵影響，而欲探究「教師的何種作為能夠有效提升學生學習成就」以及「這些作為如何去培養、增能」的關聯（大江耕太郎、大根田賴尚，2023）。因此除了學生背景問卷外，埼玉學調也包含「非認知能力」、「學習策略」、「自主、互動式深度學習」（主体的・対話的で深い学び，該詞英文為 Active learning，以下簡稱 AL），以及「班級經營」等面向之調查。施測的構念與細項，如表 6 呈現。

2015 年埼玉學調首次實施以來，埼玉縣便重視官學合作，將學調的結果委託予慶應大學綜合政策學部的中室牧子教授團隊進行分析，團隊應用教育經濟學之理論與統計方法探討影響學生學力的成因，結果如圖 8 所示。

表 6

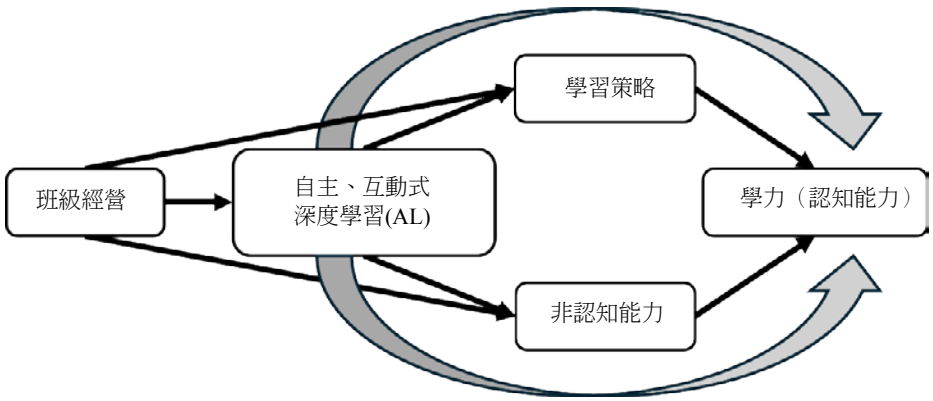
埼玉學調其他面向調查

構念	子項目與題例
1. 非認知能力	1-1 自我控制（自制心、Self-control）
	■我在家裡或是學校會發脾氣，打人或砸東西
	1-2 自我效能（自己効力、Self-efficacy）
	■我認為自己能夠弄懂課本上最難的問題
	1-3 盡責性（勤勉性、Diligence）
	■完成作業之後，我會仔細檢查是否正確無誤
	1-4 恆毅力（やり抜く力、Grit）
	■即使是失敗也不會喪失動力
2. 學習策略	2-1 彈性策略（柔軟的方略）
	■學習的時候會去考慮學習方法是否適合自己
	2-2 計畫能力（プランニング方略）
	■學習之前會先規劃好計畫
	2-3 執行功能（作業方略）
	■學習過程中遇到重點會筆記記下來
	2-4 互動能力（人的リソース方略）
	■學習中遇上困難時會向朋友尋求協助
	2-5 認知策略（認知的方略）
	■會使用自己能夠理解的詞彙來理解學習內容
	2-6 努力觀（努力調整方略）
	■學習過程中遇上困難會想要放棄，或是只學簡單的部分
3.AL	■透過聽取朋友的想法，能夠讓我更瞭解文章內容的表現方法 ■解決一道題目之後，我能夠記住所學的知識
4. 班級經營	■學校的老師是否傾聽並幫助解決你的煩惱？ ■老師是否在課堂上或測試中，耐心教導你直到你理解錯誤或不懂的部分？

註：修改自埼玉県教育委員会（2020）。

圖 8

埼玉學調因果分析示意圖



註：修改自埼玉県教育委員会（2024）。

該研究應用路徑分析，發現四個變數之間存在可預測性，研究結果得出四項結論：（1）數據證明，相對於單方面的填鴨式灌輸，通過 AL 能夠更有效地提高學習能力；（2）僅強調主動、對話式深度學習尚且不足，還要確實改變學生的學習策略與非認知能力；（3）AL 以及人際關係、信任關係能夠有效提高學生學習策略與非認知能力，這不僅對學力提升有所助益，更能夠影響一位學子的整體性人格培養，同時也呼應學校教育目的；（4）班級經營同 AL 一樣重要，它能夠拉近師生之間的關係，以提高教學效能（大根田賴尚，2019）。因果模型不僅能用以探究影響學力的諸多因素，也能使許多行動者的行為、態度「具體化」為能夠評估的構念。所以，研究結果雖然合乎既定印象，但也著實應證了理論的有效性，並能鼓舞教學現場的教育工作者（大根田賴尚，2019）。

（三）回饋教學現場的問題解決

前面提到，埼玉學調重視個體的成長進程，基於這項原則執行長期追蹤。埼玉縣教育部門根據調查結果研製「學習成長紀錄卡」（由於原文「コバトンのびのびシート」相對不易翻譯為中文，此處根據該工具功能譯之）。如圖 9 所示，每位學生的學習成長紀錄卡標記該年（2019）學調的施測結果，包含學科成績以及各項非認知能力指標，且包含前一年（2018）的施測結果。

圖 9

埼玉學調學習成長紀錄卡範本

學科	國語	年度	正確率	等級	長幅	聽、說、寫	讀	傳統文學	
		2020	63.3	6-B	3	25.0	88.9	58.8	
	算數	年度	正確率	等級	長幅	算數	測量	圖形	數量關係
		2020	40.3	5-C	0	36.4	42.9	28.6	57.1

背景問卷調查	項目	內容		2019	變化
	AL	「自主、互動式深度學習」之達成度		3.0	0.0
	學習方法	彈性策略	學習的時候會去考慮學習方法是否適合自己	3.1	0.3
		計畫能力	學習之前會先規劃好計畫	2.0	-0.2
		執行功能	學習過程中遇到重點會筆記記下來	2.8	0.3
		互動能力	學習中遇上困難時會向朋友尋求協助	1.6	0.3
		認知策略	會使用自己能夠理解的詞彙來理解學習內容	2.6	0.1
		努力觀	學習過程中遇上困難會想要放棄，或只學簡單的部分	3.1	0.3
	非認知能力	恆毅力	對事物抱持熱情，並能夠堅持不懈地投入的能力	2.7	0.2

註：擷取並翻譯自埼玉県教育委員会（2023）。

學習成長紀錄卡能夠提供教師、家長與學生參閱，藉此分析學生的學習樣態，以發揮 DDI 之效。圖 9 中的虛線圓圈指涉學生在「圖形」的得分欠佳，彰顯教師能夠協助學生精進的方向。除供 DDI 應用，埼玉學調的成果也被期望能發揮 EBPM 之效，以協助教學方法、教育行政上的改善。大江耕太郎與大根田賴尚（2023）指出，資料解讀須搭配教學現場實際狀況與教師主觀知覺，在此可參閱表 7 當中的區塊組合。

表 8 上區塊方面，做為一名教師，在教學過程中可以藉由觀察或感受來判斷一位學生或是班級整體的學習狀況，另一方面，埼玉學調的數據能呈現出具體的學生、班級學力成長幅度。倆倆搭配能夠產生出教師主觀知覺與資料一致（情況 1、4）以及不一致（情況 2、3）的狀況。這四種情況能用以對應下方區塊，比方說，當教師主觀認為學生學習狀況良好，且數據上確實也有顯著成長趨勢時（情況 1），則應採方案 1，亦即保持既有的教學方法。當教師認為學生學習反應不良，數據也顯示其表現退步時，則採行方案 4，即教師需檢討並改善現有教學方式。而發生情

況 2 和 3 時，則應採方案 2 和 3，亦即需與其他教師討論，共同商議與確認學生的學習狀態，再進一步擬定教學策略。

表 7
學調資料應用於教學之範例

	數據呈現學力提升	數據未呈現學力提升
主觀感知學力提升	情況 1	情況 2
主觀感知學力未提升	情況 3	情況 4
	搭配	
	去年度教學為適切的	去年度教學需要改善
本年度可維持既有教學	方案 1	方案 2
本年度需更改教學方案	方案 3	方案 4

註：修改自大江耕太郎與大根田賴尚（2023，頁 227）。

整體而言，埼玉學調能夠比擬為一種精密的基礎建設，相比於前面談到的 S-P 表重視於 DDI 的應用，以及沖繩案例偏重 EBPM 的實踐，鑒於埼玉學調在測驗技術上的精良，可同時供政策制定者與教學工作者兩方進行應用。

三、沖繩縣案例：參照社會之具體實踐

「參照社會」（Reference Societies）一詞主要源自比較教育、政策借用領域。在 PISA、TIMSS 等國際評比興起之後，參照社會的概念開始廣泛被應用於詮釋跨越傳統地理疆域，達到跨國政策學習之現象（Baek & Nordin, 2023; Sellar & Lingard, 2013; Volante & Klinger, 2023），舉例來說，芬蘭曾在 PISA 中表現優異，因此吸引各地學者前去考察。若將全國學調視為實踐於日本國內的微型 PISA，也會產生參照社會的現象，例如在全國學調中表現亮眼的縣市，將會成為探勘的目標，當地的教學方式也將會被擴大檢視，包含秋田縣及其下轄的東成瀨村，以及石川、福井縣等。在此以沖繩縣為例，探討全國學調如何成為一種媒介，促進日本國內縣市之間的政策學習行動。

筆者者根據國立教育政策研究所釋出的各年度報告，繪製出 2007 年首次全國學調實施以來至 2018 年，沖繩縣的成績變化如圖 10（因後續試題不再區分 A、B 類，故只擷取至 2018 年）。

圖 10

沖繩縣於全國學調各科之歷居排名

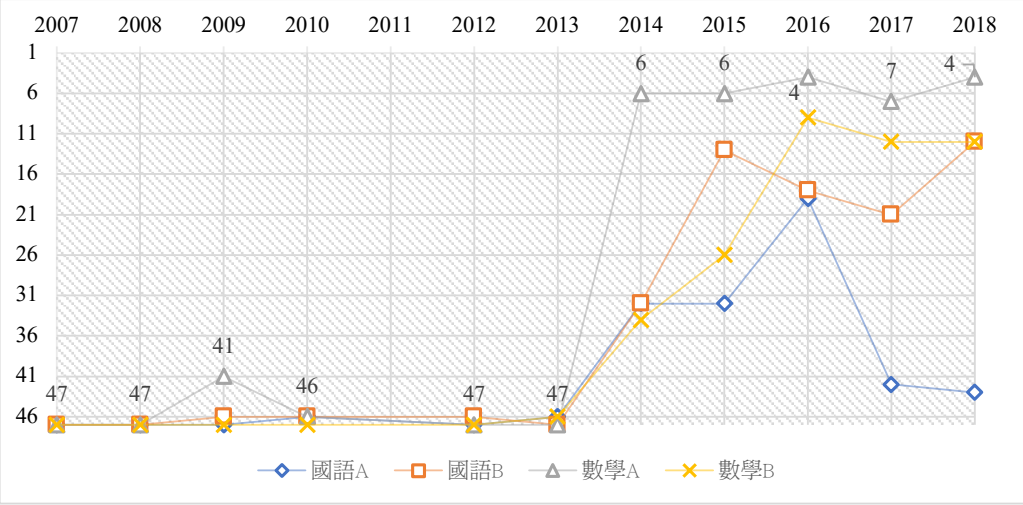


圖 10 顯示沖繩縣在全國學調四個施測項目上於 47 個自治體中的排名變化，可以發現至 2014 年以前，沖繩縣的各項表現基本上都是敬陪末座，然而在 2014 年開始有所提升，其中數學 A 更是從最後一名躍居全國第六，此後各項目雖然變動幅度各異，但除國語 A 之外，其餘都位居前段。究竟沖繩縣教育部門當局採取何種行動，以令縣內學生表現能大幅進步？

首先關於沖繩地區學生學習長期以來居於全國平均之下的原因，可從歷史、文化和經濟角度分析。根據諸見里明（2020）的描述，近代以前，沖繩先民自成琉球王國，擁有其固有語言、文化，與日本國互不隸屬，在 19 世紀日本明治維新，實施近代化改革之後，琉球王國才併入日本領土，成為沖繩縣。儘管明治維新積極推動日本現代化，然位處邊陲的沖繩仍大多以一級產業為生，相對貧困。二次世界大戰末期，沖繩縣是日本唯一遭到美軍登陸作戰的地區，許多基礎設施也因而被摧毀，日後該地更被美軍接管，直至 1972 年才回歸於日本政府，由此顯示歷史與文化因素造就沖繩與本島之間所存在的扞格。此外由於土地貧瘠，產業發展不易，沖繩縣的經濟狀況相較不佳。相關報導也顯示，沖繩縣有日本境內最高的失業率與離婚率。這些因素造成縣內高風險家庭比率居高不下，連帶影響學子學習狀況。

除了社會結構因素直接影響教育品質外，學力落後的原因也與縣內教育方針之規劃有關。曾任沖繩縣北中城小學校長的目取真康司（2019）表示，事實上沖繩

縣教育委員會對於教學精進的投入不遺餘力，但偏重在基礎學科的培養，不合時代趨勢，特別受到國際評比的影響後，學習指導要領開始更加重視思考力、判斷力與表現力等多層次素養的發展。如此一來，在全國學調進行之前，沖繩縣各校對於新型學習指導要領的認識相當模糊，在學調執行之後，各方教育工作者才開始積極思考改進之道。

2007 年全國學調首次實施，沖繩縣的成績位居末位，當時的教育委員會研擬對應措施，包含設立「學力向上推進本部會議」，希望透過學調結果來提出改善之策，然成效有限。隔年沖繩縣在學調上的排名仍無提升，此結果令民眾相當不滿，除了撥打批評電話責難當局，他們甚至要求教育委員會成員盡速請辭以示負責，此狀況足以顯示當時教育瀰漫緊繃的氛圍。其後，由曾任沖繩縣教育委員會會長的諸見里明為首，開始推動一系列的教育改革方案，這些改革主要分為對外與對內兩部分，前者為積極向其他於學調表現優良的自治體進行考察，後者則是針對縣內表現優異的學校進行訪查，讓該校的教學方法得以推廣。

（一）向外：與秋田縣的交流

沖繩縣當局選擇與當時在學調表現中名列前茅的秋田縣締成合作，促成人事交流，從 2008 年開始就派員前往秋田觀摩（伊藤綾，2008），同時也邀請秋田縣的教師擔當訪問教師，在沖繩地區任教，並邀請其提出教改上之建言。對那些前往觀摩的教師來說，可以學習到秋田縣學校內部如何訂立教學規範，並經營良好學習氛圍，還可細緻觀察到學生對於環境整潔的重視，和教師緊密的關係連結（琉球新報社，2013）。除教學面外，行政人員也建立專屬的學習評估網站，縣內工作者能夠善用此工具分析學調（目取真康司，2019）。值得注意的是，除了日本本土，秋田縣名列前茅的原因也廣泛受到海外國家吸引，部分原因包含該地學校善用探究教學法以及特殊的筆記形式，或是整個村莊對教育的重視度等（主婦の友社，2019）。

（二）向內：縣內訪查與校長領導之角色

作為教育長官，諸見里明仔細觀察縣內各校表現，發現部分學校學調成績高於全國平均，透過訪查發現，校長領導風格與學力表現高度相關。為了能夠改變校長的領導作為，他偕同委員會成員進行一系列訪視，針對教學、領導方案，進行觀察並給予指導建議（目取真康司，2019；諸見里明，2020）。經歷一系列改革，沖繩縣的學調成績於 2014 年顯著提升，此結果給予教育工作者相當大的鼓舞。

上述三個案例大致從正向角度說明學調的益處，然而實施過程中也面臨不少

質疑的聲浪。舉例而言，全國學調的營運需要投入巨額預算，根據文部科學省（2011）之公布，2007 首次全國學調的總花費預算約為 62 億日幣。如此高額的支出每被不少民眾和教育工作者質疑成效不彰，實際改善教學的程度模糊。此外，資料運用的負面事件也經常發生。這些負面事件部分屬於違反資訊保密原則，例如，日本參議院（2009）提出的質問主意書（類似臺灣的質詢意見）中即指出，雖明令不許公開市、町、村與學校的成績，然違反此規定的事件頻傳，部分負面事件屬於違反學調設計初衷，例如，有些補習機構透過大量練習來提升學調表現，以及部分地區將學調視為全國學術競賽，希冀藉此提高各校危機意識。部分案例屬於不當利用學調成果，如長年於學調中墊底的大阪市，市長為了改善教育品質，希望以學調成果做為衡量教師補助津貼與學校預算的指標，此舉招到來自官方與學界的質疑（產經新聞，2018）。諸如此類負面狀況，皆反映全國範圍的學習調查須更謹慎考量其設計與實踐。

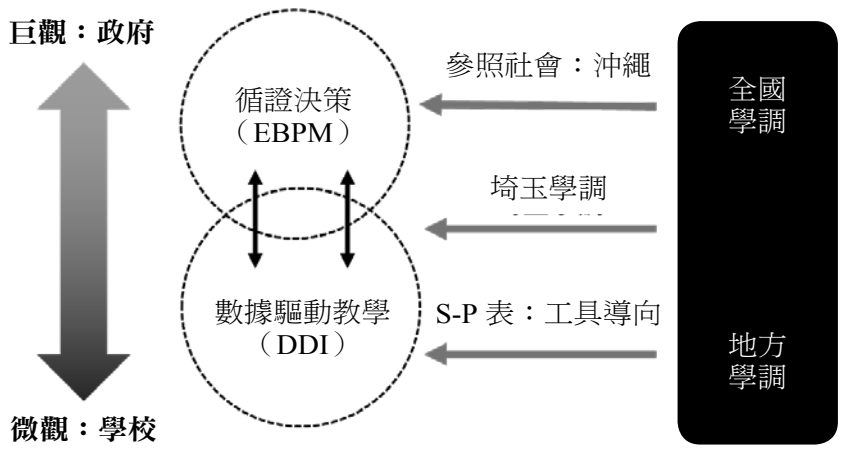
綜觀日本學調發展的演變，可以發現新版與舊版的學調在功能優劣上有所相似，兩者在設計上皆希望透過數據蒐集，得以釐清教育政策實施成效與檢測教育品質，或是力求透過競爭形式，提高學校動能，本文實質上呈現了日本全國與地方學調在提供 EBPM 與數據驅動教學的優勢。然也存在共同的缺點，教育工作者會將學調過度視為績效評估的工具，進而激化不良競爭，違背了評量設計初衷。此外，標準化評量也將可能限制了教育多樣與自主性。總之，透過日本學調的發展的雙面性，能夠作為臺灣在推動相關工作的前車之鑑。

肆、結論、討論與建議

本研究主要貢獻在於利用日本全國與地方學調的應用來釐清 EBPM 與 DDI 兩種當前新興概念的具體實踐方法，我們利用圖 11 可以更為清晰呈現學術概念與實體物件、工具、方法的相對位置。若將應用範疇區分為政府、學校層級，那麼 EBPM 比較偏向鉅觀政策導向，而 DDI 則聚焦在微觀的教學現場，然而兩者互有交疊之處，EBPM 的思維可能會驅使教師重視數據識讀，而 DDI 的成果將可能成為政策執行的參照。雖區分為全國與地方兩類，但是本文將日本學調視為一種 NALSA 的集合名詞。沖繩案例偏重政策層面，主要探討評量能夠間接發揮參照社會的作用，細緻來說呼應了 Weiss 七種證據使用模型中知識導向與問題解決模型，這是由於學調的成果能做為評估一個地區教育品質良窳的指標，同時點出了問題的

起點與教育政策的目標。琦玉學調則因其工具精良性，能同時詮釋 EBPM 與 DDI 兩種概念，因果模型建立之後能驅使政策層面更為重視非認知能力教學，而縱貫評量的優勢能夠令教師更能因應學生特質調整教學。最後，許多地區所應用的 S-P 表，能夠當成協助轉譯數據，進行意義生產的工具，進而有效將資料轉化為資訊，供教學者參考。

圖 11
研究結果示意圖



關於日本學調對於我國教育領域推動 EBPM 與數據驅動的借鏡。正如緒論所言，我國尚缺乏全國層級的教育調查評估工具，如會考、分科測驗等高風險測驗屬「test」之範疇，無法實質回饋至教學者端，且應缺乏學校、教師端之調查，因而無法評估影響學力的因子。較為接近「assessment」性質的包含各地方局處辦理之學力檢測，或如國教署推動之「學習扶助科技化評量」，但如與日本學調案例相比，這些評估能夠發揮的 EBPM 成效可能仍相當有限（梁榮仁，2019）。此外，雖然我國在國際評比 PISA、TIMSS 表現優異，這些大型調查確實也可能提供政策上的回饋，但因匿名性、採抽樣而非普查、施測週期長（3 至 4 年一次）等限制，也較具有 NALSA 的功能。事實上，諸如國家教育研究院發展之「臺灣學生成就長期追蹤評量計畫」（Taiwan Assessment of Student Achievement: Longitudinal Study, TASAL），或是目前積極推動的數位教學平臺（如：均一平台、因才網等），皆可參酌本文發現。

然而，臺灣是否需要像是日本學調的工具，也可以值得深思，畢竟就日本的案例來看，學調的應用性也非完全正向，甚至有可能引發出更多問題。然就具體建議上，臺灣教育部門若欲設計類似評估工具，可在研發之前深思目的與性能。例如，川口俊明（2020）認為評量的功能可分為「指導導向」與「政策導向」，前者似於形成性評量的概念，意指評量本身的目的主要在於協助教師進行認知診斷，而後者則是將測驗作為判斷政策實施有效性的工具。茲以表 8 說明。

表 8
評量工具的兩種功能

功能	指導導向	政策導向
舉例	藉由測驗找出大部分學生不易撰寫的漢字，強化該方面教學，或是觀察哪些學生在漢字撰寫上遭逢困難，增進其輔導	欲判斷縮小班級規模是否能夠有效提升每一位學生的學力。如果該政策無法有效提高學力，就不必投入增聘教師的成本來實施小班教學。
形式	普查	抽樣
目標	使所有學生都達到預訂標準	能夠有效鑑別政策實施前後平均成就的差異
回饋	提供學生即時的測驗結果和學習診斷	提供教育當局調查的分析結果，供其決策參考
其他	不太需要關注試題本身以外的調查細項與分析指標。	重視群體而非個體的描述統計量（如班級平均數）及其誤差。 需要調查學生背景變項（如家庭社經背景、學校學習環境）。

參照上表標準，兩種類型之測驗結果的分析分別對應的是不同領域的人才需求。倘若重視政策導向，則須結合社會學、經濟學、政策等專業人士；若重視實踐回饋，則需與各學科領域學者、現場教師協作。值得注意的是，全國學調被批評的其中原因即在於投入大量的經費，卻無法完整服膺任何一種測驗的功能（川口俊明，2020）

本文所提供的日本學調案例能作為臺灣在推動 EBPM 或是 DDI 上的參考，然在本文較少提及卻需要補充的是，資料的「公開性」是所有行動的前提。據筆者淺見，臺灣學界不乏各類能夠將數據有效轉譯為嚴謹證據、資訊的專家，然可能囿於公部門對於各類資料的管控較為嚴格，使其較為難以被研究者或是學術機關使用並進行分析，因此在提倡 EBPM 或 DDI、DDDM 的同時，也鼓勵產、官、學之間形塑數據共享、信任的文化。

誌謝

本研究承蒙國家科學及技術委員會「補助博士生赴國外研究計畫」（千里馬計畫）（MOST 112-2917-I-003-005）經費補助，特此致謝。初稿並蒙匿名審查者提供寶貴修改建議，謹此一併感謝。

參考文獻

久永隆一（2023 年 12 月 5 日）。**PISA ってなに？全国学力調査が復活したきっかけ**。2024 年 7 月 2 日，取自 https://www.asahi.com/articles/ASRD44TWPRD1UTIL031.html

大江耕太郎、大根田頼尚（2023）。**現場で役立つ！教育データ活用術**。日本評論社。

大根田頼尚（2019）。一人ひとりを伸ばす教育をめざして。載於千々布敏弥（主編），**学力がぐんぐん上がる急上昇県のひみつ**（頁 76-85）。教育開発研究所。

小杉聡（2023）。全国学力調査問題の統一でテストは変わったのか――小学校国語科の問題を中心に―。**日本国際情報学会誌『国際情報研究』，20**（1），27-38。

川口俊明（2020）。**全国学力テストはなぜ失敗したのか**。岩波書店。

今井康晴（2008）。我が国の「教育内容の現代化」におけるブルーナー教育理論の受容に関する一考察。**明星大学教育学研究紀要，23**，97-106。

内閣官房（2022）。**EBPM ガイドブック**。2024 年 7 月 17 日，取自 https://www.gyoukaku.go.jp/ebpm/img/guidebook1.0_221107.pdf

内閣府（2020）。**海外におけるＥＢＰＭの事例とエビデンス集作成について**。2024 年 7 月 17 日，取自 https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/ab1/20201027/shiryou5.pdf

文部科学省（2018）。**平成 30 年度実施予定の都道府県・指定都市による独自の学力調査について（小学校）**。2024 年 6 月 17 日，取自 https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/sonota/detail/__icsFiles/

afieldfile/2018/06/26/1406389_1.pdf

文部科学省（2019a）。**教育政策におけるＥＢＰＭの強化**。2024 年 7 月 17 日，取自 https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/wg7/20191203/shiryou1.pdf

文部科学省（2019b）。**平成 30 年度全国学力・学習状況調査活用事例集（2）**。2024 年 8 月 15 日，取自 https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/__icsFiles/afieldfile/2019/03/25/1413828_02.pdf

文部科学省（2019c）。**平成 30 年度全国学力・学習状況調査活用事例集（4）**。2024 年 8 月 15 日，取自 https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/__icsFiles/afieldfile/2019/03/25/1413828_04.pdf

文部科学省（2024a）。**令和 6 年度全国学力・学習状況調査報告書【小学校／国語】**。2024 年 7 月 19 日，取自 https://www.nier.go.jp/24chousakekkahoukoku/report/data/24plang.pdf

文部科学省（2024b）。**令和 7 年度以降の全国学力・学習状況調査（悉皆調査）の CBT での実施について**。2024 年 8 月 7 日，取自 https://www.mext.go.jp/content/20240430-mxt_chousa02-000035736-1.pdf

主婦の友社（2019）。**「学力日本一！」秋田県東成瀬村のすごい学習法**。

北野秋男（2015）。全国の市町村教育委員会による「学力調査」の実施状況：学力評価体制の実態と構造。**教育學雜誌**，**51**，17-31。

目取真康司（2019）。全国学力調査最下位の衝撃から 10 年。載於千々布敏弥（主編），**学力がぐんぐん上がる急上昇県のひみつ**（頁 40-49）。教育開発研究所。

伊藤綾（2008 年 11 月 3 日）。**沖縄、秋田に学ぶ学力調査で明暗、人事交流を検討**。2024 年 8 月 5 日，取自 https://www.asahi.com/edu/student/news/TKY200811030064.html

足立区教育委員会（2024）。**足立区における学力調査結果の活用について**。2024 年 8 月 15 日，取自 https://www.mext.go.jp/b_menu/shingi/chousa/shotou/112/shiryo/__icsFiles/afieldfile/2016/05/23/1367805_1_1.pdf

参議院（2009）。**全国学力・学習状況調査の有効性等に関する質問主意書**。2024 年 8 月 9 日，取自 https://www.sangiin.go.jp/japanese/joho1/kousei/syuisyo/171/syuh/s171150.htm

国立教育政策研究所（2014）。**平成 26 年度全国学力・学習状況調査の調査問題・正答例・解説資料について**。2024 年 6 月 17 日，取自 https://www.nier.go.jp/14chousa/14chousa.htm

国立教育政策研究所（2025）。**教育データサイエンスセンター**。2025 年 3 月 25 日，取自 https://www.nier.go.jp/04_kenkyu_annai/div12-data-sci.html

松下佳代（2014）。PISA リテラシーを飼いならす―グローバルな機能的リテラシーとナショナルな教育内容――。**教育学研究**，**81**（2），14-27。

松下佳代（2017）。学力とは教育学の観点から。**日本労働研究雑誌**，**680**，55-57。

浦岸英雄（2010）。全国学力テストはなぜ実施されたのか。**園田学園女子大学論文集**，**40**，27-39。

琉球新報社（2013 年 11 月 5 日）。**県挙げた環境参考に教員 27 人学力先進地、秋田視察**。2024 年 8 月 2 日，取自 https://ryukyushimpo.jp/news/prentry-214832.html

埼玉県庁（2025）。**統計からみた埼玉県のすがた 2025 年版**。2025 年 3 月 21 日，取自 https://www.pref.saitama.lg.jp/a0206/a360/kennosugata2025.html

埼玉県教育委員会（2020）。**埼玉県の学力向上施策の概要**。2024 年 6 月 23 日，取自 https://www5.cao.go.jp/keizai-shimon/kaigi/special/reform/ab1/20201027/shiryou3.pdf

埼玉県教育委員会（2023）。**令和 2 年度埼玉県学力・学習状況調査報告書（埼玉県学力・学習状況調査を活用した実践事例）**。2024 年 6 月 19 日，取自 https://www.pref.saitama.lg.jp/f2214/gakutyou/r2tuika.html

埼玉県教育委員会（2024）。**令和5年度埼玉県学力・学習状況調査報告書（第1章）**。2024 年 6 月 28 日，取自 https://www.pref.saitama.lg.jp/f2214/r5gakutyouhoukokusyo/r5gakutyouhoukokusyo.html

望田研吾（2005）。学力向上には何が必要なのか。**教育と医学**，**53**（2），192-198。

産経新聞（2018 年 8 月 11 日）。**大阪市長「学力テスト結果、給与反映」方針に広がる波紋、渦巻く賛否**。2024 年 8 月 9 日，取自 https://www.sankei.com/article/20180811-HLXTL3I3PJML7GL4KOM56V3ILE/

梁榮仁（2019）。新北市國小學生能力檢測現況探討與省思。**臺灣教育評論月刊，8**（9），頁 92-99。

高橋徹（2022）。戦後の体育改革に影響を与えたデューイの教育論の射程：矮小化からの解放と議論の前提の再構築。**体育学研究，67**，9-23。

富山県中学校教育研究会（2024）。**SP 表分析について**。2024 年 8 月 15 日，取自 https://www.ktkk.ico.bz/upfile/spp.pdf

間渕泰尚（2019）。学習者から見た「ゆとり教育」に対する評価。**神戸親和女子大学児童教育学研究，38**，279-291。

遠藤貴広（2010）。日本の場合——PISA の受け止め方に見る学校の能力観の多様性——。載於松下佳代（主編），**〈新しい能力〉は教育を変えるか——力・リテラシー・コンピテンシー——**（頁 181-202）。ミネルヴァ書房。

諸見里明（2020）。**学力テスト全国最下位からの脱出——沖縄県学力向上の取組み**。学事。

Baek, C., & Nordin, A. (2023). Understanding implicit reference societies in education policy. *Journal of Education Policy*, *39*(5), 736-754. https://doi.org/10.1080/02680939.2023.2282456

Bambrick-Santoyo, P. (2010). *Driven by data: A practical guide to improve instruction*. John Wiley & Sons.

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *47*(2), 108-121. https://doi.org/10.1111/1467-8527.00106

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking connections between research and practice in education: A conceptual framework. *Educational Researcher*, *47*(4), 235-245. https://doi.org/10.3102/0013189X18761042

Filderman, M. J., Toste, J. R., Didion, L., & Peng, P. (2022). Data literacy training for K-12 teachers: A meta-analysis of the effects on teacher outcomes. *Remedial and Special Education*, *43*(5), 328-343. https://doi.org/10.1177/07419325211054208

Hargreaves, D. (2007). Teaching as a research-based profession: Possibilities and prospects (The Teacher Training Agency Lecture 1996). In H. Daniels, H. Lauder, & J. Porter (Eds.), *Educational research and evidence-based practice* (pp. 3-17). Routledge.

Howlett, M. (2009). Policy analytical capacity and evidence-based policy-making: Lessons

from Canada. *Canadian Public Administration*, *52*(2), 153-175. https://doi.org/10.1111/j.1754-7121.2009.00070_1.x

Kaspi, S., & Venkatraman, S. (2023). Data-driven decision-making (DDDM) for higher education assessments: A case study. *Systems*, *11*(6), 306. https://doi.org/10.3390/systems11060306

Lai, M. K., & Schildkamp, K. (2013). Data-based decision making: An overview. In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (pp. 9–21). Springer. https://doi.org/10.1007/978-94-007-4816-3_2

Lee, J., Lee, J.-S., & Ellis, N. (2023). Public opinion about national large-scale student assessment: A case of NAPLAN. *Educational Assessment*, *28*(3), 137-155. https://doi.org/10.1080/10627197.2023.2215977

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, *34*(2), 209-231. https://doi.org/10.3102/0162373711431604

Lingard, B., Martino, W., Rezai-Rashti, G., & Sellar, S. (2015). *Globalizing educational accountabilities*. Routledge.

Mandinach, E., Honey, M., & Light, D. (2006, April 9). *A theoretical framework for data-driven decision making* [Paper presentation]. The Annual Meeting of AERA, San Francisco, United State. https://cct.edc.org/sites/cct.edc.org/files/publications/DataFrame_AERA06.pdf

Marsh, J. A., & Farrell, C. C. (2015). How leaders can support teachers with data-driven decision making: A framework for understanding capacity building. *Educational Management Administration & Leadership*, *43*(2), 269-289. https://doi.org/10.1177/1741143214537229

Morozumi, A., & Tanaka, R. (2023). *School accountability and student achievement: Neighboring schools matter*. https://www.rieti.go.jp/jp/publications/dp/23e004.pdf

Nilsson, M., Jordan, A., Turnpenny, J., Hertin, J., Nykvist, B., & Russel, D. (2008). The use and non-use of policy appraisal tools in public policy making: An analysis of three European countries and the European Union. *Policy Sciences*, *41*(4), 335-355. https://

doi.org/10.1007/s11077-008-9071-1

Ninomiya, S. (2019). The impact of PISA and the interrelation and development of assessment policy and assessment theory in Japan. *Assessment in Education: Principles, Policy & Practice*, *26*(1), 91-110. https://doi.org/10.1080/0969594X.2016.1261795

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using evidence: How research can inform public services*. Bristol University Press. https://doi.org/10.2307/j.ctt9qgwt1

Organisation for Economic Co-operation and Development. (2023). *Education at a Glance 2023: OECD Indicators*.

Poortman, C. L., & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education*, *60*, 425-433. https://doi.org/10.1016/j.tate.2016.06.010

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163-180. https://doi.org/10.1177/0165551506070706

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration*, *80*(1), 1-22. https://doi.org/10.1111/1467-9299.00292

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, *61*(3), 257-273. https://doi.org/10.1080/00131881.2019.1625716

Schildkamp, K., & Visscher, A. (2010). The utilisation of a school self-evaluation instrument. *Educational Studies*, *36*(4), 371-389. https://doi.org/10.1080/03055690903424741

Sellar, S., & Lingard, B. (2013). Looking East: Shanghai, PISA 2009 and the reconstitution of reference societies in the global education policy field. *Comparative Education*, *49*(4), 464-485. https://doi.org/10.1080/03050068.2013.770943

Sonobe, T., & Kurosaki, T. (2024). Implementing empirical economics into policymaking. In K. Otsuka, T. Kurosaki, Y. Sawada, & T. Sonobe (Eds.), *Next-generation of empirical research in economics* (pp. 35-56). Springer. https://doi.org/10.1007/978-981-97-1887-0_3

Takayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, *44*(4), 387-407. https://doi.org/10.1080/03050060802481413

Takayama, K., & Lingard, B. (2021). How to achieve a 'revolution': Assembling the subnational, national and global in the formation of a new, 'scientific' assessment in

Japan. *Globalisation, Societies and Education*, *19*(2), 228-244. https://doi.org/10.1080/14767724.2021.1878016

Teodoro, A. (2022). *Critical perspectives on PISA as a means of global governance: Risks, limitations, and humanistic alternatives*. Routledge.

Thompson, G. (2013). NAPLAN, MySchool and accountability: Teacher perceptions of the effects of testing. *International Education Journal*: *Comparative Perspectives*, *12*(2), 62-84.

Volante, L., & Klinger, D. A. (2023). PISA, global reference societies, and policy borrowing: The promises and pitfalls of 'academic resilience'. *Policy Futures in Education*, *21*(7), 755-764. https://doi.org/10.1177/14782103211069002

Volante, L., & Mattei, P. (2024). The politicization of PISA in evidence-based policy discourses. *Policy Futures in Education*, *22*(8), 1554-1569. https://doi.org/10.1177/14782103241227309

Weiss, C. H. (1979). The many meanings of research utilization. *Public Administration Review*, *39*(5), 426-431. https://doi.org/10.2307/3109916

2024 年 8 月 19 日收件
2025 年 3 月 25 日第一次修正回覆
2025 年 4 月 24 日通過初審
2025 年 4 月 28 日第二次修正回覆
2025 年 5 月 29 日第三次修正回覆
2025 年 7 月 17 日第四次修正回覆
2025 年 8 月 22 日複審通過